

MEETING DOCUMENT

Wadden Sea Board (WSB 38)

06 October 2022
Tønder, Denmark



Agenda Item:	5.3 Trilateral Monitoring and Assessment Programme
Subject:	TMAP data handling proposal
Document No.:	WSB 38/5.3/2
Date:	20 September 2022
Submitted by:	TG-MA / EG-Data / CWSS

Access to and documented quality of Wadden Sea data is an essential component of TMAP, as is the provision of monitoring data available for trilateral assessment and as a basis for visualization to reach out with information on the ecological assessment of the Wadden Sea World Heritage.

During its 36th meeting, the WSB agreed that TG-MA, together with EG-Data, would coordinate a study to investigate the necessary resources for implementing the different data handling scenarios, including the option to incorporate additional data, based on the outcomes of a common data workshop in January 2022, together with international experts in this field. Due to the expected high efforts, CWSS commissioned a consultancy to provide such study in close cooperation with EG-Data. For gaining a broader perspective, a consortium of external experts from the Danish Environmental Portal and the International Council for the Exploration of the Sea, ICES, who also attended the January data processing workshop, contributed to the process in an advisory capacity. During its development, the report turned out to be a comprehensive basis for any future implementation of a data handling system. This proposal is aiming at setting the scene for the revitalization of the information and data streams within TMAP under consideration of latest technologies and acceptable efforts for the regional data centers whose internal structures should remain largely untouched. Nevertheless, any activities can only be launched with an initial one-off effort to install necessary structures but are aiming at the same time for limited efforts in the long run, based on a high level of automation.

The preliminary data handling report on TMAP by Brockmann Consult is available on request and not attached due to its extent.

Proposal: The meeting is invited to agree on the outlined procedure towards a future TMAP data handling as outline in the chapter 4. “The way forward”.

1. Background

TG-MA together with EG-Data have explored several scenarios for organizing the trilateral data handling under TMAP. The formerly submitted scenario document considered a stepwise approach towards the formerly agreed sophisticated outcome, with varying efforts to be made for interim results (data handling arrangements). Apart from a disorganized data handling as current, unsustainable practice, four incremental scenarios were developed. To be more precise on the technical implementation and resulting needs for resources, a data handling study was commissioned to Brockmann Consult, which has a long-time record of involvement in trilateral data handling and therefore the necessary insights to be able to perform such study within the limited time frame. The study was performed in close cooperation with CWSS. As a supporting activity, members of EG-Data were requested to complete questionnaires on the regional TMAP data arrangements to understand the different set-ups as basis for finding a common solution. Focus was on the technical implementation details under each scenario but also on the expected resources needed, but not on reaching out for a detailed system architecture even though major elements were captured (see also Annex I). This document is summarizing the findings of the report in a condensed way and outlines activities necessary for implementing a future trilateral data handling.

Any approach developed in the report and taken up in this document shall fulfil the following requirements:

- no changes or only minor changes will be applied to the national/regional databases,
- available regional infrastructure shall be considered,
- any basic data quality control happens at national/regional level,
- the FAIR data principles shall be followed (FAIR Principles - GO FAIR (go-fair.org).” (Ref.)¹

2. Initial findings

In contrast to former assumptions to deal with the different scenarios separately, it turned out that any evolution toward a sophisticated solution would require a step-by-step approach. With the default of leaving the regional data centers largely untouched and to limit access to the already available information from these sources, it became clear that the initial, but probably most important, step will be to compile a minimum set of metadata describing the technical specifications of the data sources required for largely automated data processing (“TMAP data catalogue”). Assembling this information for all parameters is an unavoidable first step to guarantee a minimum of availability of common data sets for the Wadden Sea region. This will result in initially increased efforts to access detailed information on the currently 182 parameters to be provided once by the regional data centers. A test series with e.g. 10 parameters could be initially performed as a proof of concept. A realistic timeline for such activities will need to be agreed on with the involved organizations. Nevertheless, this approach guarantees a high flexibility in adding or removing parameters if considered as necessary. The most prominent features and consequences expected under each scenario are listed in table 1. Schematic diagrams of major architecture elements of each scenario are displayed in Annex I.

3. TMAP data handling scenarios

3.1 Towards scenario 1

Based on the availability of data source information described above, a specific website shall be created allowing all users access to the web-based data sources, or alternatively direct them to the responsible people, with only limited efforts. This access will be provided via a “central access point” represented by one or multiple webpage/s. To this point, all information would be provided unharmonized over the regions, apart from parameters which are standardized by their nature e.g. species numbers. This information level

¹FAIR = Findability, Accessibility, Interoperability, and Reuse

will allow at least QSR authors, but also CWSS, to access trilateral data for further processing. Any harmonization effort is on the side of the client (QSR author) associated with a correspondingly high level of effort (costs). This strategy is reflecting the arrangements formerly described under “scenario 1” and was also considered in EG-Data and TG-MA as an absolute minimum approach and a transition phase. The limitation is that providing and displaying Wadden Sea wide information in a common format is not possible due to limited comparability for a significant number of parameters.

3.2 Towards scenario 2

From this point onwards, the regional data centers will be a lot less affected regarding mandatory efforts. In order to achieve a certain level of data harmonization, the available data sets will need to undergo transformations by web-based tools to be set-up at the central data center (“initial harmonization toolbox”) which still need to be developed for purpose. The necessary development processes would be assigned to EG-Data together with CWSS, external advice, and specific projects to be commissioned to third parties. While a status corresponding to scenario 1 will be achieved, the timeline for these activities can be set in a realistic way to continuously upgrade the harmonization level and data quality with the aim to limit the efforts for authors and also to create simple or sophisticated web-based interactive maps. A lot of TMAP data sources are already available via the internet for the purpose of visualisation but the level of harmonisation is rather limited.

3.3 Towards scenarios 3 and 4

During the process of creating the report, it turned out that the steps needed to achieve any features described for the scenarios 3 and 4 were not considered as different enough to separate these scenarios. Again, the additional effort for the regional data centers to reach these scenarios is rather limited as the underlying data sets were extracted from the regions at an earlier stage. A partly harmonized data set for each parameter could serve as basis for any kind of web interface (e.g. the Cadenza software agreed to be used in former approaches). The options will only be limited by resources provided, but could, however, be implemented at any time once a stable data delivery and processing will be achieved. Scenario 3 would stand for a simple visualization and scenario 4 for an extended display. Both options need to be further defined.

Table 1: Expected consequences of the different TMAP data handling scenarios.

„Status quo“			
	Scenario I	Scenario II	Scenarios III and IV
Access	<ul style="list-style-type: none"> central data access point 	<ul style="list-style-type: none"> central data access point 	<ul style="list-style-type: none"> central data access point
Costs	<ul style="list-style-type: none"> limited costs for data collection, but potentially high costs for data processing for QSR authors 	<ul style="list-style-type: none"> implementation costs for technical data harmonization 	<ul style="list-style-type: none"> implementation costs for data harmonization & visualization
Data quality	<ul style="list-style-type: none"> control on data sources different levels of quality, consistency, and harmonization 	<ul style="list-style-type: none"> control on data sources further control on quality/consistency of harmonization 	<ul style="list-style-type: none"> control on data sources further control on quality/consistency of harmonization
Outreach	<ul style="list-style-type: none"> limited outreach potential beyond QSR 	<ul style="list-style-type: none"> limited outreach potential beyond QSR 	<ul style="list-style-type: none"> S3: simple data visualization (Geoviewer can also serve as a data portal) -> enhanced outreach potential S 4: extended data visualization -> high outreach potential

4. Expected efforts

The entire set-up described up to this point is based on a sustainable and easy-to-maintain infrastructure. The task to come up with an assumption on the resources needed under each scenario was completed to the extent possible. Since it was difficult to assess the infrastructure and Full-time equivalent (FTEs) needed at this early stage in the process, the information provided can only be considered as an approximation and can only be detailed, specified and finalized by EG-Data. Once the requirements for Scenario 1 are met, it is foreseen to limit the efforts for the regional data centers to regular maintenance and updating the parameter information. The data handling report is listing the estimated efforts for the different TMAP data handling scenario implementations in detail in chapter 4 “Components, tasks and efforts to implement the scenarios”, already considering a stepwise approach. The essence of the assumption is listed in this document in table 2 and 3.

Table 2: Estimated required efforts under each scenario (s) for contractor, regional data centers (RDC) and central data center (CDC). The values represent workdays and are grouped by the scenarios and by the duration of the commitments (initial, i. e. once, or long-term, i. e. per year).

Scenario	Contractor	Each RDC	CDC
S1 (initial)	55 (65)	5	0
(long-term)	0	5	6
S2 (initial)	90	21	0
(long-term)	0	10 ²	21
S3 (initial)	35	1	0
(long-term)	0	1	10
Sum (initial)	180	26	0
(long-term)	Maintenance contract	15	37

Table 3: Table 24: TMAP data handling scenarios: Required software at the regional and central data centers (RDC / CDC).

Scenario	Component	Software	Data center
1	Database(s)	Database software	RDC
	Download service	Download link, OGC web service, API	RDC
	Central TMAP database	Database software	CDC
	Central access point	Web application providing dynamic web pages for data download	CDC
		Tomcat application server ³⁾	CDC
2	Optional: Metadata service (if metadata are not provided by OGC services)	Metadata catalogue software, perhaps provided by other agencies, e. g. the "National georegister" in the Netherlands or "Geodateninfrastruktur Deutschland" and "Marine data infrastructure-DE" in Germany.	RDC
	OGC web services (WFS, WMS)	OGC WFS web services, e. g. Geoserver or Deegree	CDC
3	Simple web GIS	Webclient (e.g. Leaflet or OpenLayers) and Tomcat (see scenario 1)	CDC
4	Extended web GIS	Option 1: Dedicated TMAP-specific webGIS application and Tomcat (see scenario 1) Option 2: Commercial configurable web GIS software (e.g. Cadenza)	CDC

²⁾ 5 units will be required for long-term maintenance and 5 units for mid-term set-up of supplemental web services.

³⁾ This server will run the web application required for dynamic webpages and optionally the web GIS application for scenarios 3 and 4.

5. The way forward

TG-MA and EG-Data welcomed the comprehensive TMAP data handling report from Brockmann Consult, as well as the valuable input from the Danish Environmental Portal and ICES. Both groups consider a stepwise approach, with the implementation of scenario 1 as a starting point, as a feasible option but also as a major step forward also towards further elaborated scenarios which comply with the agreed purpose of TMAP (a set up similar to scenario 4). Considering the current dissatisfying situation (“status quo”) and the defaults of the regional data centres (regional structures remain largely untouched as such), the envisaged data availability via a web-based central access point would be a smart step forward, even if harmonization efforts for some parameters would not be taken into account at that point. For further progress, it is advisable to start working on scenario 2 in parallel with a subset of parameters, so that initial experience with the intended data harmonisation can be gained and an appropriate workflow can be developed. Members pointed out that the further development towards more sophisticated solutions should be considered from the beginning and should be a crucial part of the tasks of TG-MA and EG-Data during the period to come, also since an enhanced outreach of TMAP, with its supporting and safeguarding function for the Wadden Sea World Heritage, was one focus of the German presidency.

Schematic diagrams of the technical features of the different TMAP data handling scenarios

Table 1: Alphabetical glossary of technical terms mainly used in ANNEX I.

Term	Explanation / Definition
API	Application Programming Interface
Central Access Point	Website compiling and listing data and metadata sources for all TMAP parameters. Data download via weblink.
Central data web service	Web services set-up in central data centers, providing access to harmonised and validated data
Data catalogue	contains all technical specifications of the available TMAP data sources in a pre-defined structured way - all characteristics to access and process the data sources are required
Harvester	Regional data will be requested and assembled in an automatized way
Harmonization Toolbox	Web service that automatically harmonizes harvested data sets in a pre-conditioned TMAP format. Requirements of the service need to be developed and overlooked by CWSS / EG-Data / TG-MA. Level of harmonization dependent on data quality and resources invested.
Validator	routines to check value ranges and codes
Regional data centers	Organizational unit responsible for regional data provision
WFS	Web Feature Service - service providing geodata (i. e. geometries and attribute data)
WMS	Web Mapping Service - service providing images, showing styled geodata

“STATUS QUO”

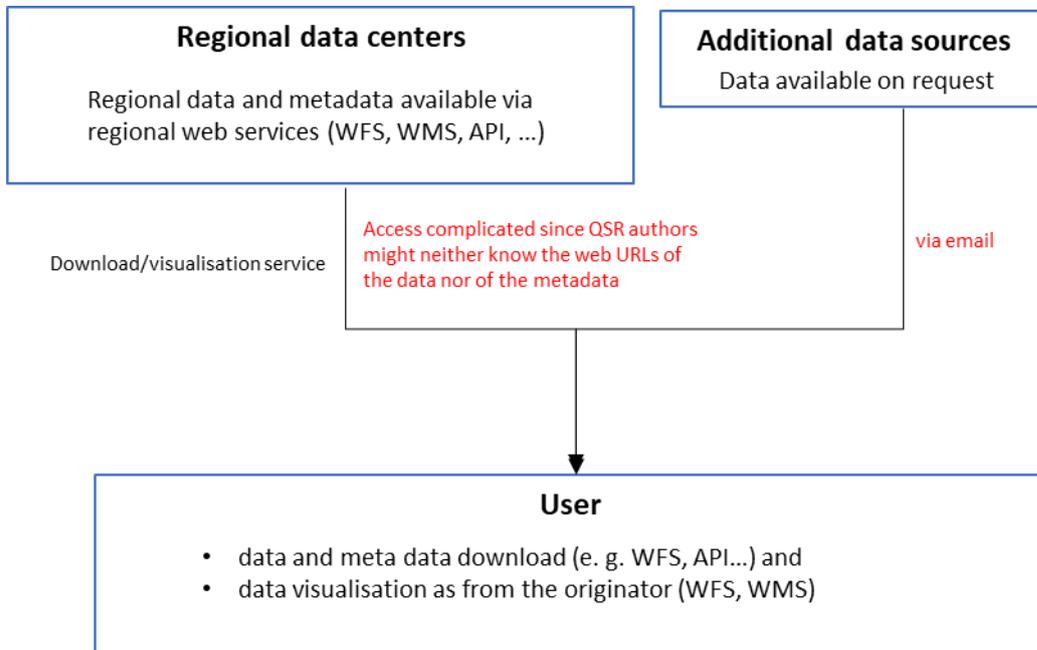


Fig. 1: Schematic overview of the current situation of the trilateral data handling (“Status Quo”)

SCENARIO 1

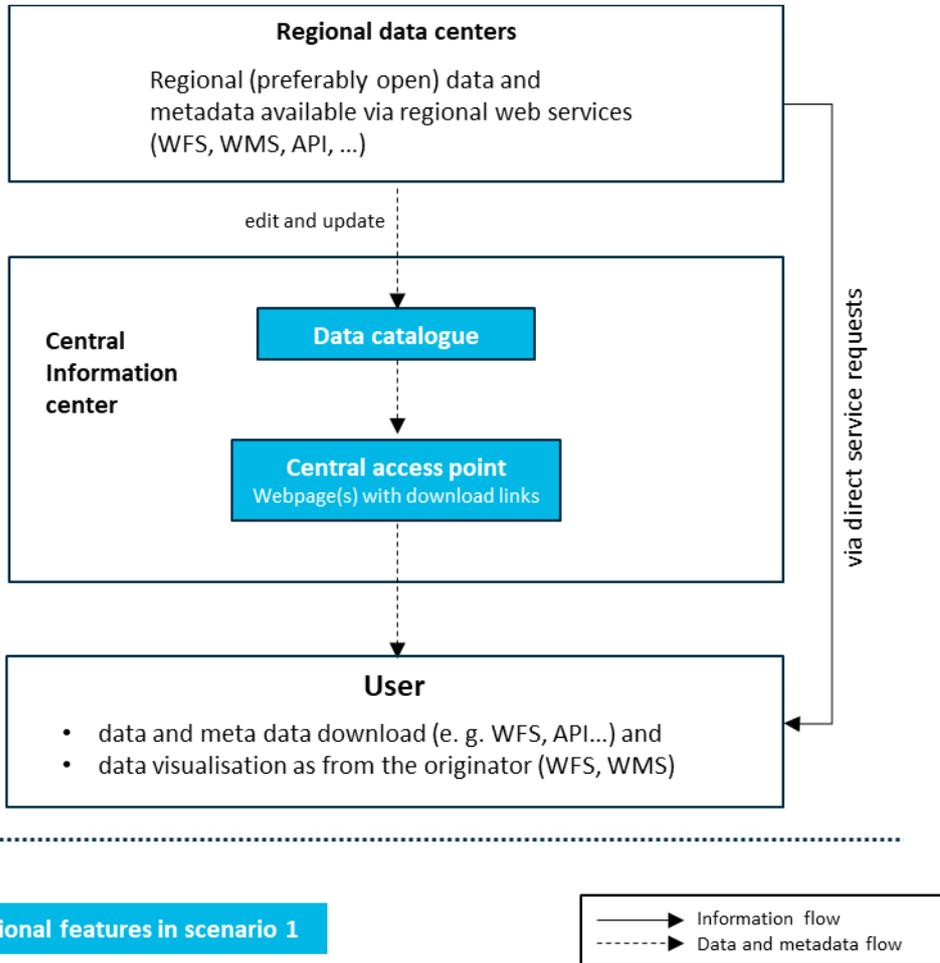
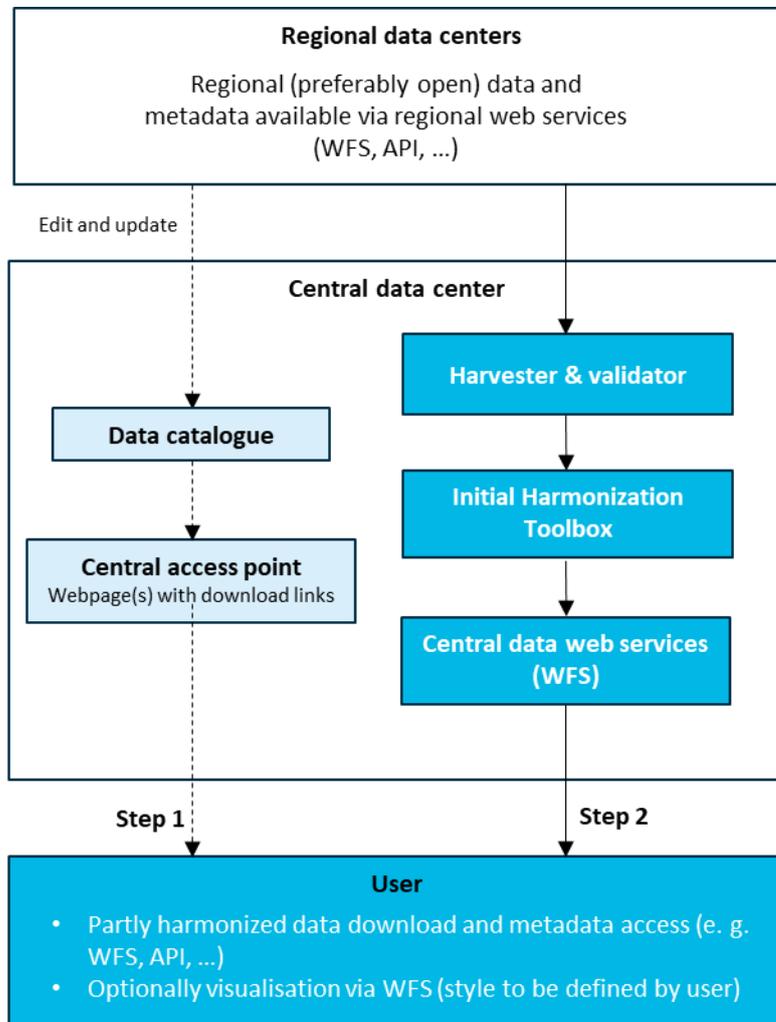


Fig.2: Schematic overview of data handling arrangements reflecting scenario 1.

SCENARIO 2



Additional features in scenario 2

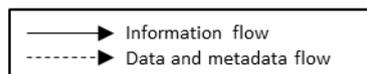
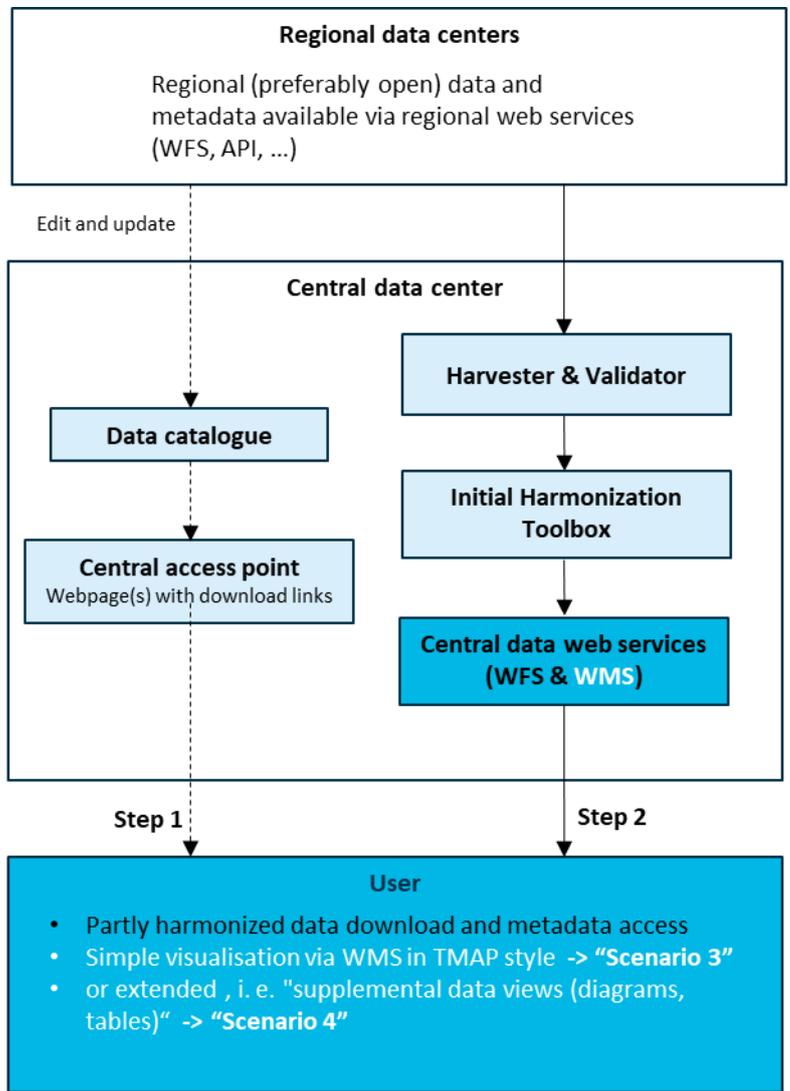


Fig.3: Schematic overview of data handling arrangements reflecting scenario 2.

SCENARIO 3 & 4



Additional features in scenarios 3 and 4

—▶ Information flow

- - -▶ Data and metadata flow

Fig.4: Schematic overview of data handling arrangements reflecting scenario 3 & 4